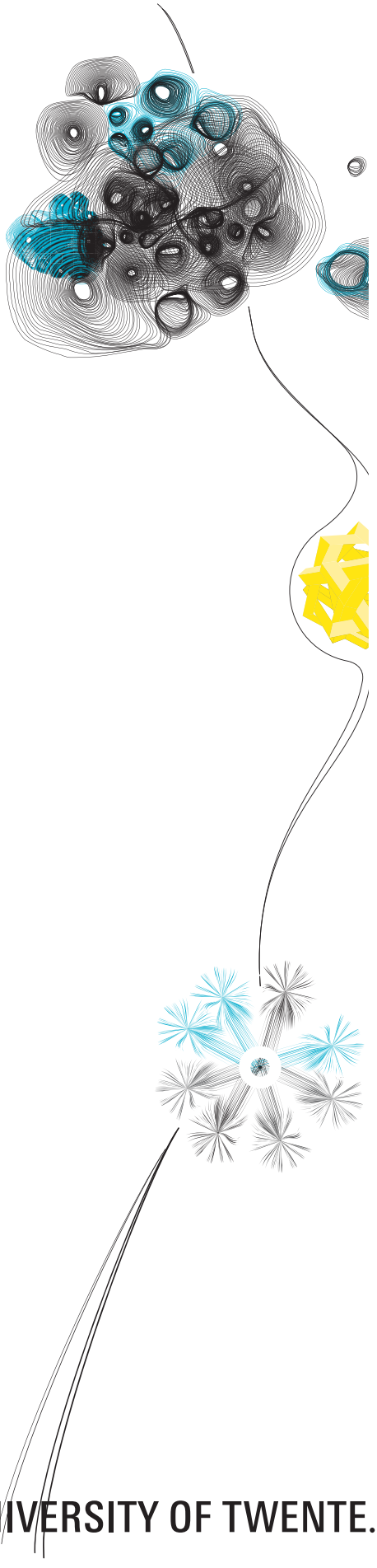

USING ACTIVITY RECOGNITION TO IMPROVE HEART RATE MONITORING ACCURACY

FINAL PAPER

By:
Maarten Thoonen
mawthoonen@gmail.com

Faculty of Electrical Engineering, Mathematics & Computer Science
Biomedical Signals & Systems Group
Programme: Embedded Systems

Supervised by:
Y. Wang PhD
Prof. Dr. Ir P.H. Veltink
Dr. F. R. Halfwerk
Dr. Ir. R.W. van Delden

UNIVERSITY OF TWENTE.

MAY 12, 2022

***Abstract*—Heart rate monitoring in continuous automatic vital signs monitoring often produces false alarms, because alarm thresholds are static and do not take context into account. Alarm management could be improved by taking into account what an expected heart rate would be. This work attemps to create a data-driven model of the heart rate reponse to various activities. An experiment was conducted which resulted in a dataset with 9 participants. In this experiment, participants wore IMUs (Inertial Measurement Unit, a type of movement sensor) and a dry electrode ECG recorder and were asked to perform Activities of Daily Living (ADL). The dataset was used to create a data-driven model that predicts the heart rate from the current activity. To do so, a Support Vector Machine (SVM) classifier with Radial Basis Function (RBF) kernel is used to perform Human Activity Recognition (HAR). A *k*-NN regressor is used to perform heart rate prediction based on the predicted activity, activity intensity and activity duration. Five-fold cross validation was used to evaluate system performance. The HAR classifier had a median accuracy of 87%, with a minimum of 82% and a maximum of 92%. The heart rate prediction algorithm had a median absolute error of median 3.82 BPM, minimum 2.94 BPM and maximum 4.79 BPM. This is deemed an acceptable result for a preliminary system. Future work includes building a more practical system to wear, creating a more general model that does not need to be individually trained and clinical validation.**

***Index Terms*—Human Activity Recognition, Machine Learning, Heart Rate Monitoring, Vital Signs**

I. Introduction

IN CURRENT care, nurses measure a patient's vital signs every eight hours. Usually, at least the heart rate, blood pressure, temperature and SpO_2 are checked [1]. This takes up a considerable portion of nurses' time. Additionally, a patient's condition can worsen substantially in eight hours. Therefore, it would be better if the vital signs are checked more often than that, as that would allow for deteriorating patients to be caught earlier [2], [3]. Continuous automatic vital sign monitoring allows constant patient monitoring, so clinical deterioration is detected on time and the required interventions can be given earlier. An additional advantage is that an automatic continuous vital sign monitoring system is usually more portable than bed-side monitors, and could therefore be more convenient in challenging areas, such as field hospitals [4]. However, automatic continuous vital sign monitoring has proven to be a difficult challenge [5]. Factors like portability, patient comfort and continuous availability all have an effect on system accuracy. Consequently, an inaccurate system would produce false alarms, leading to 'alarm fatigue'—i.e., the tendency to ignore alarms when they occur too often. Nurses ignoring alarms could have potentially lethal consequences. Nevertheless, research suggests adaptation in clinical care practices is likely if the incidence of false alarms is reduced to an acceptable level [6].

To prevent false alarms, the system should deal with problems that arise during continuous monitoring. Normally, vital signs are only measured when a patient is at rest. With continuous monitoring this is not necessarily always the case, particularly because light exercise is common in current care, as it accelerates patient recovery [7]–[9]. This exercise leads to elevated heart rates and motion artefacts. The system should not raise alarms due to expected higher heart rates during exercise or due to motion artefacts. A possible solution is to let a monitoring system take patient activity into account. Patient activity could be automatically recognized using Human Activity Recognition (HAR) methods.

HAR has been an active field of research for decades. The field can be split up into three main categories: HAR using a network of simple sensors, HAR using cameras and HAR using movement sensors. A simple sensor network works with a number of switches and motion sensors embedded in doorways and objects. A central system then tracks where people are in the environment and with which objects they are interacting [10], [11]. Computer vision can be used to detect activities with very high precision and accuracy [12], [13]. Recent research suggests that this is also possible with anonymized, privacy-preserving video data [14]. Activity recognition using accelerometers goes back a long time [15], and really took off after Bao & Intille showed that Machine Learning could be applied for activity recognition from accelerometer data [16]. Since then, various research has been published trying different techniques, sensor and settings. Some try different Machine Learning techniques with wearable movement sensors [17]–[20]. This can also be the movement sensor from a smartphone [21], [22]. HAR has also been tried with patients, with good results [23], [24]. Recently the focus has been on deep learning [25], which can also be used to perform HAR on datasets containing many different types of sensors, a practice called sensor fusion [26]–[28].

This work focuses on improvement of heart rate monitoring in particular. Several mechanisms influence the human heart rate. The signals that drive the heart muscles are generated by the sinus node in the heart. The sinus node responds directly to chemical and mechanical stimuli, but also indirectly to a wide range of stimuli through the sympathetic and parasympathetic (via the vagus nerve) nervous systems [29], [30]. Those nervous systems are influenced by various factors—e.g., exercise [31], stress [32], circadian rhythm and homeostasis. Other mechanisms that influence the heart rate are the breathing rate through the Respiratory Sinus Arrhythmia (RSA) and the baroreflex, which is the response of the heart rate to changes in blood pressure.

Wearables specifically for measuring vital signs are already on the market. Recent research indicates they are ready for clinical use [33], [34]. However, there are some remaining challenges in decision making. Currently, hospitals have to manually set alarm thresholds in these systems. These thresholds are static and potentially subjective. By creating a model of the heart rate for various levels and types of activity, this work attempts to create a system that compares a measured heart rate with a predicted heart rate. Comparing the measured and predicted heart rate values could assist in alarm decision making. Because of the many factors that influence the human heart rate, it is difficult to create a comprehensive model that accurately predicts the heart rate corresponding to a certain activity. Physiology-based mathematical models are complicated, computationally intensive and model only the heart rate in limited situations [35], [36]. Machine Learning could aid in overcoming these limitations [37]. Therefore, this work leverages Machine Learning to create a data-driven

model.

II. Method

A. Experiment

The experiment was ethically approved by the Ethics Committee Computer & Information Science of the University of Twente (reference number RP 2021-188). Based on prior experience and comparable literature [38]–[48], 14 participants were included. All participants are assigned a random integer number to differentiate between them in the Results section. To exclude any effects from adverse health, only healthy participants were considered. All participants were recruited from around the University of Twente.

The experiment was conducted in the eHealth House. This is a lab at the University of Twente that resembles natural living conditions. Facilities include a living room, kitchen, bedroom and bathroom. In the bedroom are a chair, double bed and a closet. The living room and kitchen are one large room, where a seating area for lounging, a kitchen table with chairs and a large kitchen counter can be found. There is also a control room that is raised approximately 1 meter above the floor of the rest of the lab, reachable by stairs. The lab is equipped with cameras and microphones to monitor and record experiments. Monitoring can be done through a web interface or from the control room. Additionally, there is a microphone in the control room to talk to experiment participants. Therefore, the researcher does not need to be in the same room as the participants. This is useful to avoid artificial influences and during the COVID-19 pandemic to prevent contamination. A floorplan of the lab can be found in Fig. 1. The bathroom and briefing/debriefing room are not used during the experiment.

Six main activity classes were considered: standing, sitting, lying, walking, walking up or down stairs and cycling. These activities were based on comparable literature [20], [23] and movements cardiac patients perform on the postoperative ward in a hospital[1]. The experiment was split up into three main parts: controlled, free and cycling. The controlled part consisted of standing, sitting, lying and walking in a controlled manner, at a fixed location and for a known length of time. Walking was done at three different speeds: slow, normal and fast. The exact speed was open for interpretation by the experiment participant, so he or she walked in a natural rhythm. The path was chosen to be as long as possible, and is indicated in Fig. 1 by the dotted line with arrows. Its total length is approximately 12 meters. Participants were instructed to sit on the edge of the bed or on a chair and to lie in a supine position, as lying prone or on the side might have interfered with the sensors.

During the free part of the experiment, participants were given a more general task, like ‘make a cup of tea’ or ‘fetch an object from another room’. Participants were also instructed to walk up and down the stairs to the control room at three different speeds: slow, normal and fast. Again, the participants chose the exact speed.

[1]The author spend a morning with the physiotherapists at the postoperative ward of the Thorax Centrum Twente, the cardiology center of the Medisch Spectrum Twente hospital in Enschede, The Netherlands.

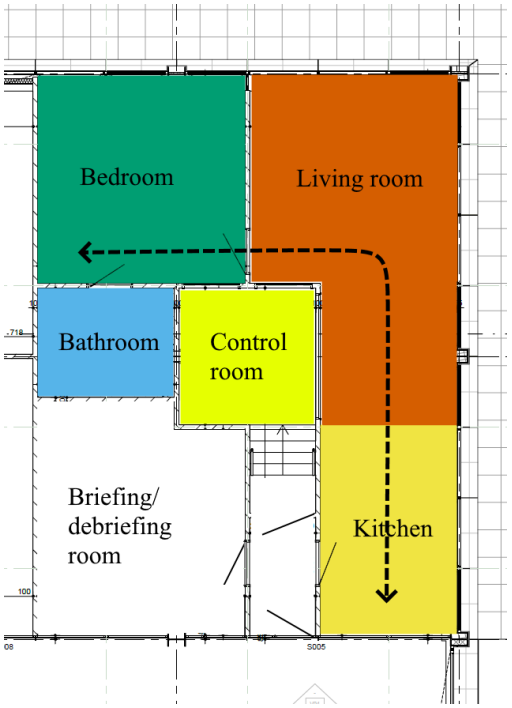

Figure 1. Floorplan of the eHealth House. In the bedroom are a chair, a closet and a double bed. In the living room is a seating area with a low table, couch and comfortable chair. The kitchen has a dinner table with four chairs and a long kitchen counter. The briefing/debriefing room has a large meeting table. In the bathroom is a shower, sink and toilet. The control room has equipment for monitoring the experiment. The dotted line with arrows indicates the walking path during the experiment. This image is modified version of the floor map received from the eHealth House contact person, with permission.

The experiment was ended by 10 minutes of cycling. The cycling consisted of three parts: three minutes of warming up at a moderate intensity, six minutes of cycling at a high power level and one minute of cooling down at a low power level. The exact intensity (power level in watt of the bicycle) of each part was different for each participant. Initial power levels and the heart rate target are based on the American College of Sports Medicine’s Guidelines for Exercise Testing and Prescription [49]. Those guidelines specify initial power levels based on gender and physical condition. A heart rate of 140 BPM was the target for the high power part. Participants cycled at a cadence of 50 rpm (rotation per minute). The researcher monitored the heart rate during cycling and adjusted the power level accordingly. After cycling, participants filled in a short survey while seated. Therefore, activities such as sitting, standing and walking also occur during this part. The questions of the survey were the participant’s age, length and weight, how often they exercise, whether they think anything affected their heart rate, whether they think anything affected their movements and any other remarks. The experiment protocol was based on Activities of Daily Living (ADL), similar studies [50] and activities physiotherapists use in a hospital setting. The complete experiment protocol can be found in Appendix A.

We applied five IMUs (Intertial Measurement Units) for physical activity recognition and one ECG (electrocardiogram)

Figure 2. Photo of the Zephyr BioHarnass and five Xsens IMUs (circled in red) as worn during the experiment. The resolution of this figure is low, as it was taken from a video recording. This participant is left-handed.

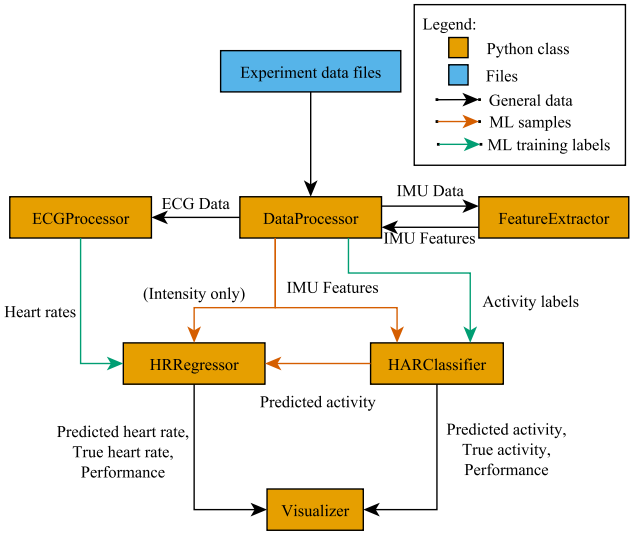

Figure 3. Workflow diagram of the data processing pipeline. ML = Machine Learning, ECG = electrocardiography, HR = Heart Rate, HAR = Human Activity Recognition.

sensor as the gold standard test for heart rate detection. The IMUs are of the type Xsens MTw Awinda [51]. These measure linear acceleration (in m/s^2) in three directions and rotational velocity (in °/s) in three directions. Data collection was done wirelessly using a base station connected to a laptop computer. A Zephyr BioHarness [52] was used for ECG recording. This device records a one-lead ECG at 250 Hz using dry electrodes, which are more comfortable for participants than normal ECG electrodes. ECG data collection was done wirelessly with a Bluetooth connection to a laptop computer. The sensor positions were chosen based on positions that yielded high accuracies in literature [20] (chest, upper arm and upper leg) and positions that are convenient for wearables (lower arm/wrist and the hip, as devices can be attached to a belt [53]). The five IMUs were attached to the chest, upper arm, lower arm, hip and upper leg. The arm sensors were attached to the participants' dominant arm, and the leg sensor to the right leg. The hip sensor was attached to the side of the hip on the same side as the upper leg sensor. A photo of all sensors attached to an experiment participant can be found in Fig. 2. This participant is left-handed, a right-handed participant would have a mirrored setup for the arm sensors.

B. Data annotation

The data was labelled by the researcher by watching back the recordings of the experiment using a custom labelling application. A short explanation of this application can be found in Appendix B. All labels were synchronized to both the IMU and ECG data by careful visual inspection of the ECG and IMU data for recognizable movement influence on the data and finding the corresponding activity label. Most labels were based on the video recording, but for walking stairs the audio recording was used, as this activity occurred outside of view from the cameras. The data was split into blocks of five seconds each and each block had an overlap of 50% with the previous block. This window was chosen as a tradeoff between activity length and the rate at which the heart rate changes. The 50% overlap ensured there was enough resolution to consider short activities, whereas the window length of five seconds ensured there were enough heart beats to calculate a BPM value from. Each data block was assigned a label based on the activity label that corresponded to the data in the middle of the data block. For cases where many short activities followed each other, we made the assumption that the activity in the middle of the data block is most relevant for the entire data block.

C. Data processing & Analysis

All data analysis was done in Python, using the Pandas [54], [55], SciPy [56] and NumPy [57] packages. A full overview of the data processing pipeline can be found in Fig. 3. The following sections elaborate on this pipeline.

1) IMU data: In case there were missing values in the IMU data due to connection issues, they were filled in using linear interpolation. Feature selection was done by reasoning what the expected signals were. When an experiment participant is stationary, all accelerometers have a constant acceleration due to gravity and all gyroscopes have a constant velocity of 0. When the participant is moving, the accelerometer signals change due to shifting positions and the gyroscopes respond in a certain way. A summary of how each sensor responds to each activity can be found in Appendix D. The conclusion of the analysis was that 24 features are sufficient to distinguish between all activities. All features can be found in Table I. These come from the chest sensor and either the upper leg or the hip sensor. Each sensor has six signals: linear acceleration in the x-, y- and z-axis (Acc X, Y and Z in the table) and rotational velocity around the x-, y- and z-axis (Gyr X, Y and Z in the table). Both the average value of and the variation in the signals were necessary to describe all activities, as both average position values due to gravity and dynamics resulting from activity have to be considered. Therefore, for each signal

Table I
OUTLINE OF THE IMU DATA FEATURES.

Sensor	Signal	Features
Chest	Acc X	Mean, Standard Deviation
	Acc Y	Mean, Standard Deviation
	Acc Z	Mean, Standard Deviation
	Gyr X	Mean, Standard Deviation
	Gyr Y	Mean, Standard Deviation
	Gyr Z	Mean, Standard Deviation
Upper leg or Hip	Acc X	Mean, Standard Deviation
	Acc Y	Mean, Standard Deviation
	Acc Z	Mean, Standard Deviation
	Gyr X	Mean, Standard Deviation
	Gyr Y	Mean, Standard Deviation
	Gyr Z	Mean, Standard Deviation

Acc X, Y, Z = accelerometer signal in the x, y and z axes.
Gyr X, Y, Z = gyroscope signal around the x, y and z axes.

the mean and standard deviation were calculated. All features were normalized to have zero mean and unit variance as this is required for many machine learning estimators from SciKit-Learn [58].

2) ECG data: For heart rate detection, the R peak of the QRS complexes in the ECG signal was used as the fiducial point for beat detection. The detection algorithm is based on the paper by Pan & Tompkins [59]. The original algorithm was modified in several ways to better fit the needs of this research, the signal from the BioHarness and to take advantage of modern hard- and software. The algorithm was verified with the MIT-BIH Arrhythmia Database [60] and the European ST-T database [61].

The first change compared to the original algorithm is the band pass filter. It has a different passband and is implemented in a different way, to better fit the noise characteristics of the ECG signal from the BioHarness and because modern hardware allows for better filters. The ECG signal was band pass filtered using an Infinite Impulse Responds (IIR) filter consisting of three second order filter sections. The passband ranged from 15 Hz to 30 Hz. This allowed the R peaks to pass, while movement artifacts and power line noise were removed. Then, the difference between all samples was taken and the resulting signal was squared. These steps ensure all peaks are positive and the sharp R peak slopes are emphasized. A moving average filter over 25 samples was then employed to ensure that the rising and falling edges of the R peak are combined into one large peak. The original moving average filter is 30 samples long, but a 25-sample filter improved noise rejection performance based on visual inspection. Longer windows caused R peaks to merge with noise peaks, which caused the time at which the beat was detected to shift. The signal was capped to a maximum value, as larger values are always caused by artifacts and those outliers unjustly modified the detector thresholds.

Noisy sections of the ECG signal were identified by counting all peaks of a certain prominence within a 1-second interval. By visual inspection of the raw and preprocessed ECG signals, it was determined that if the preprocessed signal contained more than five peaks per second, that section of the signal corresponds to a noisy ECG that is not suitable for beat detection. This peak counter was not in the original Pan-Tompkins algorithm, but is an addition used for this work.

The detector works by first checking if a certain sample is a peak—i.e., the sample is a local maximum. If that was the case, it was checked whether that peak is above a certain threshold and if the last R peak is a realistic amount of time in the past (one fifth of the average of the last five R-R peak distances, or 0.3 seconds for very high heart rates, as the maximum heart rate was assumed to be 200 BPM). If those conditions were met, the peak was labelled as an R peak. If the peak was not above the R peak threshold, but the last detected R peak was more than 1.4 times the average of the last five R-R peak distances in the past, a lower threshold was used to check whether peaks are R peaks. Some thresholds were different than the original Pan-Tompkins algorithm, but they were verified with the MIT-BIH Arrhythmia Database and the European ST-T database. When a peak was classified as an R peak, the detector thresholds were updated. If a peak was classified as not an R peak, it was classified as a noise peak and the noise threshold was updated accordingly. By continuously updating those thresholds, the algorithm adapts to the ECG signal. BPM values were calculated for every detected R peak. The BPM value corresponding to a certain R peak was defined as the average of the last eight R-R peak distances. In case there were noisy sections causing R peaks to be missed, the average BPM values of the sections before and after the noisy section were used to interpolate BPM values for the noisy section.

D. Machine Learning

Machine learning was applied for two parts of this research: HAR and heart rate prediction. The Python package SciKit-Learn [62] was used to implement the machine learning algorithms. Five-fold cross validation was used to verify all machine learning algorithms. This means five different splits of the data into 80% training data and 20% testing data were created. All performance scores are the average score for the five different splits. This meant a separate validation dataset was unnecessary. This validation strategy was chosen as it is often used for datasets of this size in combination with classical machine learning methods.

1) Human Activity Recognition: The data was split into a training set and a testing set using a stratified five-fold splitter with shuffling enabled. Using a stratified split ensured that each activity class is represented equally in all subsets. Enabling shuffling ensured that not only consecutive samples were in each subset, which means all subsets contain data from the controlled, free and cycling parts of the experiment. Activity recognition was attempted for all participants separately, creating individual classifiers, and with the data aggregated from all participants, resulting in a general classifier. From a literature search it was concluded that a Support Vector Machine (SVM) or a *k*-Nearest Neighbors (*k*-NN) classifier

is most suitable for HAR using IMU signals [17]–[22], [24], in particular as there are no computing resource constraints for this exploratory study. From initial testing it became clear the SVM performed better on this dataset, so this classifier was optimized.

An SVM gives the best performance if the dataset is linearly separable. This means that sets of samples are separable by a hyperplane. In the case of a two-dimensional dataset with two sample classes, that means a straight line can be drawn between the two groups of samples in a graph depicting that dataset. The input samples for the classifier were the 24 IMU features as explained in Section II-C1, and both the hip and the upper leg sensor were tried. This dataset was initially not linearly separable, but a transformation kernel could be employed to transform the dataset into a linearly separable dataset. The best performance was obtained with a Radial Basis Function (RBF) kernel, which can be found in (1).

$$K(\boldsymbol{x}, \boldsymbol{x'}) = \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{x'}||^2}{2\sigma^2}\right) \tag{1}$$

The value of a transformed sample depends on the distance between the input sample $\boldsymbol{x}$ and a central sample $\boldsymbol{x'}$ and on a free parameter σ. The values of $\boldsymbol{x'}$ and σ were automatically found by an algorithm. The algorithm has a parameter `gamma` which determines how well the function fits the data. Lower values lead to inaccurate classifiers, but a too high value leads to overfitting. To find the best value, a grid search was used. A grid search is a systematic parameter sweep, each with a five-fold cross validation. From the results of the grid search followed that the best value for this parameter was 0.1. For some individual experiment participants different values would lead to a slight increase in performance, but the same value was chosen for all participants to improve comparability between participants. The SVM also has a regularization parameter `C`, which determined the smoothness of the decision functions. A lower value of `C` leads to smoother decision functions, whereas a higher value leads to a better fit to the training data, at the expense of longer training time. A grid search was used to decide on a value of 10.0 for the parameter `C`. The activity labels from Section II-B corresponding to the input samples were used as training and test labels.

The performance of the classifier was assessed using the f1-score. Equation (2) is the definition of the f1-score:

$$\text{f1-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{2}$$

Precision and recall are defined in (3) and (4):

$$\text{Precision} = \frac{tp}{tp + fp} \tag{3}$$

$$\text{Recall} = \frac{tp}{tp + fn} \tag{4}$$

Where tp means true positive, fp means false positive and fn means false negative. True positive means all correctly predicted labels of a certain activity class A, false positive means all instances of another activity class that are incorrectly predicted to be activity class A, and false negatives are all instances of an activity class A that are incorrectly predicted to be from another activity class. From those definitions it follows that precision means the fraction of times that a sample was correctly predicted to be class A versus the total number of samples that is predicted to be class A, correct or not. Recall means the fraction of all correctly predicted labels from activity class A versus the total number of samples from class A. Precision, recall, f1-score, tp, fp and fn all range from 0 to 1.0. A score of 0 means 0% of all predictions are described by that metric and 1.0 means 100% of the predictions is described by the metric.

2) Heart Rate Regressor: The heart rate prediction algorithm was implement using a *k*-Nearest Neighbors regressor. A *k*-NN regressor calculates the output by taking the average of the *k* most similar training samples. An advantage of a *k*-NN regressor is that it is relatively good at capturing non-linearities in low-dimensional data [63]. Additionally, because we expect that similar activities at similar intensities lead to similar heart rates in most cases, a *k*-NN regressor is a logical choice. Input data considered for the regressor are activity labels, activity intensities and activity durations. The regressor used the same five-second data window with 50% overlap as the HAR classifier. Activity intensity is defined as the average of all standard deviation values from Table I, scaled to fit in a range from 0 to 1. The choice whether the upper leg sensor or the hip sensor was used, was based on which sensor performed best for HAR. Activity duration is the amount of consecutive data windows the current activity has been going on for. The average measured heart rate during one data window was used to label each window. Both the manually created activity labels and the predicted activity labels (the output from the HAR classifier) were tried with the heart rate regressor. This way, a best-case scenario could be compared with a real-world scenario. A five-fold splitter with shuffling enabled was used to split the dataset into training and testing data. Whereas the activity classifier had only the six activity classes as possible outcomes, the regressor has an infinite number of possible BPM values as possible outcomes. Therefore, a stratified splitter cannot be used, so the data is only shuffled. A *k*-NN regressor has the parameter `n_neighbors` that determines of how many samples the average is taken. Using a grid search with five-fold cross validation, the optimal value for this parameter was determined to be 14. The regressor is trained separately for each participant.

The performance of the heart rate regressor was evaluated with the R^2-score, also known as the coefficient of determination. This metric calculates the proportion of variation in the output data that is explainable by the input variables. It is defined in (5):

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

Where $\hat{y}$ is the predicted output value, y is the true (input label) value and $\bar{y}$ is the average of all values y. The subscript $_i$ denotes a certain sample. The R^2-score ranges from 0 to 1.0, where higher is better. A score of 0 means 0% of the variation of the output is causes by the input, so it is essentially random. A score of 1.0 means that 100% of the output is explained by

Table II
TABLE CONTAINING DEMOGRAPHICS OF THE EXPERIMENT PARTICIPANTS.

	Age (years)	Length (cm)	Weight (kg)
Median	22	179	65
Minimum	19	159	45
Maximum	27	193	93

the input, meaning a perfect fit. In addition to the R^2-score, the median absolute error, mean absolute error and maximum error were used to assess the performance of the regressor. These values are expressed in BPM.

III. RESULTS

Of the 14 experiment participants, 9 resulted in a usable dataset. One was excluded because of corrupted IMU data files. The other four participants were excluded because of unusable ECG data. This was caused by a faulty sensor strap, which seemed to operate adequately while the participant was at rest, but produced excessive clipping artefacts during activity. One of the participants has less data available, as one of the sensor batteries died during cycling. Table II contains demographics of the included experiment participants. Five participants are female and four are male. Participants were asked how often they exercised, this ranged from never to seven times per week. The power level during cycling ranged from 65 Watt to 130 Watt for the high intensity part. The intensity was dependent on gender and how often participants exercised. Notable results from the survey were that one of the participants remarked that 'you are very aware of what you are doing'. Factors that participants believed influenced their heart rate were having a slight hangover, having a cold or having skipped breakfast that morning.

A. Human Activity Recognition

The best results were obtained by using the chest and upper leg sensors. Table IV contains summarized performance results from the HAR classifier. The lowest f1-score is 82%, the highest is 92% and the median is 87%. Using the hip sensor, f1-scores were lowest 78%, highest 93% and median 87%.

The full scores per activity can be found in Table III. The column 'Support' indicates the number of samples for that activity. Subscores per experiment participant and with mean heart rate values and the standard devitation of the heart rate can be found in Appendix C. It is clear that the activity 'walking stairs' in particular is underrepresented in the dataset. Due to the difficulties in labelling that activity (using the audio recording instead of video recording) and similarity to normal walking, scores are generally low for this activity. However, due to the low number of samples, this has only a limited effect on the general performance scores. This also became clear by using the hip sensor instead of the upper leg sensor: although the average scores are quite similar (hip: median 87%, min 78%, max 93% and upper leg: median 87%, min 82%, max 92%), the scores for walking stairs are consistently lower with the hip sensor than with the upper leg sensor. This is partly offset by slightly higher scores for sitting and lying. For the activity 'lying' the small amount of samples is less of a problem, as this activity is well-distinguishable from the other activities and well-labelled.

For all participants, confusion matrixes were generated to visualize the performance per activity class. Fig. 4 shows the confusion matrix for the combined dataset from all participants. Fig. 5 shows the confusion matrix for the best performing participant, which is participant 8. It is clear that most confusion occurs between the activities 'standing' and 'walking', and that 'stairs' performs relatively bad.

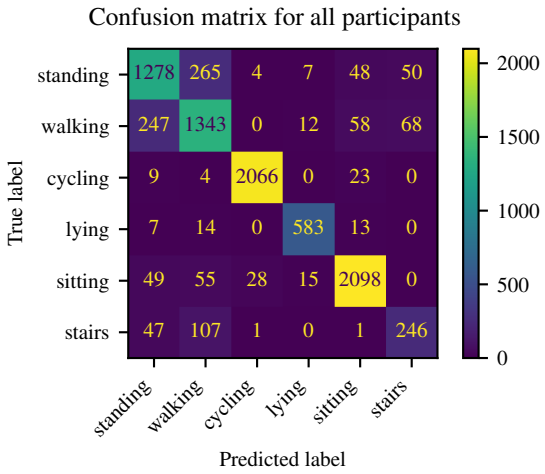

Figure 4. Confusion matrix for the combined dataset from all participants.

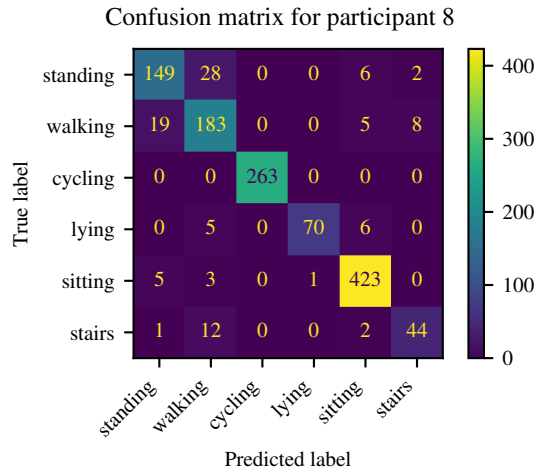

Figure 5. Confusion matrix for participant 8.

B. Heart rate prediction

The performance results for the heart rate regressor using the manually annotated activity labels can be found in Table V and in Table VI the results from using predicted activity labels can be found. The best performance was achieved by using all input features, so activity type, activity length and activity intensity. Both the chest and the upper leg sensors were used

Table III
PERFORMANCE METRICS FOR HAR PER ACTIVITY.

Activity	Precision			Recall			f1-score			Support
	median	min	max	median	min	max	median	min	max)	
Standing	0.78	0.66	0.86	0.80	0.67	0.83	0.79	0.68	0.83	1652
Walking	0.79	0.63	0.85	0.81	0.68	0.86	0.80	0.65	0.86	1728
Cycling	0.99	0.98	1.00	0.98	0.97	1.00	0.99	0.98	1.00	2102
Lying	0.99	0.88	1.00	0.86	0.71	0.98	0.92	0.83	0.99	617
Sitting	0.94	0.88	0.98	0.93	0.91	0.98	0.94	0.89	0.98	2245
Stairs	0.71	0.64	0.90	0.62	0.53	0.86	0.64	0.61	0.89	402

Table IV
TABLE CONTAINING PERFORMANCE METRICS FOR HUMAN ACTIVITY RECOGNITION.

Participant	Precision	Recall	f1-score
1	0.82	0.82	0.82
2	0.86	0.86	0.86
4	0.86	0.85	0.85
5	0.90	0.90	0.90
6	0.88	0.87	0.87
8	0.92	0.92	0.92
11	0.87	0.87	0.87
12	0.91	0.91	0.91
13	0.87	0.86	0.87
All (mean ± SD)	0.88 ± 0.028	0.87 ± 0.029	0.87 ± 0.029
Aggregate	0.87	0.87	0.87

'All' means the mean and Standard Devation of all the separate runs, while 'Aggregate' means the results for the classifier trained with data from all participants combined.

Table V
TABLE CONTAINING PERFORMANCE METRICS FOR HEART RATE PREDICTION USING MANUAL ACTIVITY LABELS.

Participant	R^2-score	Mean absolute error	Median absolute error	Max error
1	0.83	5.21	3.08	68.4
2	0.77	6.32	4.38	38.1
4	0.83	5.60	3.81	39.0
5	0.82	6.05	3.80	46.9
6	0.55	5.99	4.95	31.0
8	0.82	6.07	4.26	35.7
11	0.88	5.25	3.48	30.3
12	0.88	5.33	3.70	36.1
13	0.91	4.90	2.88	33.7
All (mean ± SD)	0.81 ± 0.10	5.63 ± 0.461	3.88 ± 0.620	39.9 ± 11.1

Table VI
TABLE CONTAINING PERFORMANCE METRICS FOR HEART RATE PREDICTION USING PREDICTED ACTIVITY LABELS.

Participant	R^2-score	Mean absolute error	Median absolute error	Max error
1	0.79	5.81	3.27	77.7
2	0.75	6.66	4.72	34.9
4	0.81	5.87	3.77	46.6
5	0.75	6.85	4.00	52.2
6	0.56	5.98	4.79	24.9
8	0.81	6.10	4.22	40.3
11	0.86	5.58	3.74	37.3
12	0.87	5.53	3.82	36.1
13	0.88	5.30	2.94	34.7
All (mean ± SD)	0.79 ± 0.094	5.96 ± 0.5483	3.97 ± 0.612	42.7 ± 14.3

to determine the activity intensity. For the best-case using the manual labels, the R^2-scores range from 55% to 91% with a median of 83%. Using the predicted labels, the R^2-scores range from 56% to 88% with a median of 81%. As expected, the regressor performs slightly worse when predicted labels are used. This is logical, as the labels are not always correct, which means the regressor uses wrong information to train and predict a heart rate. Participant 6 has a considerably lower score than the other participants. A probable cause of this is that the battery of one of the IMUs died early during cycling. Therefore, there is less heart rate data available and in the data that is available, there is less variation. Participant 1 has a large maximum error. This is caused by very sudden spikes in the measured heart rate during walking up or down stairs that are not captured by the regressor. These spikes seem to be at least partially caused by movement artefacts.

A boxplot of all errors from the regressor using predicted labels can be found in Fig. 7. The orange line shows the median error, the box covers the 25^{th} to the 75^{th} percentile and the black horizontal lines show the highest and lowest errors, excluding outliers. A value is considered an outlier if it is more than 1.5 times the interquartile range (IQR) removed from the median. As can be seen in the boxplot, the distribution of errors is concentrated around the median, but has long tails. As indicated by the fliers, most participants have many outliers. Most often, this is caused by wrong activity labels. An exception is participant 6, as there is little heart rate data for cycling and therefore the training data set does not contain many high BPM values. Also clearly visible is that for participant 1, most errors are relatively small (indicated by the horizontal black lines being close to the median), but the fliers indicate a large maximum error.

Fig. 6 shows that there is some visible correlation between the current activity and the current heart rate. We chose participant 11 for this visualization, as this participant showed a clear heart rate response to different activities. An R^2-score

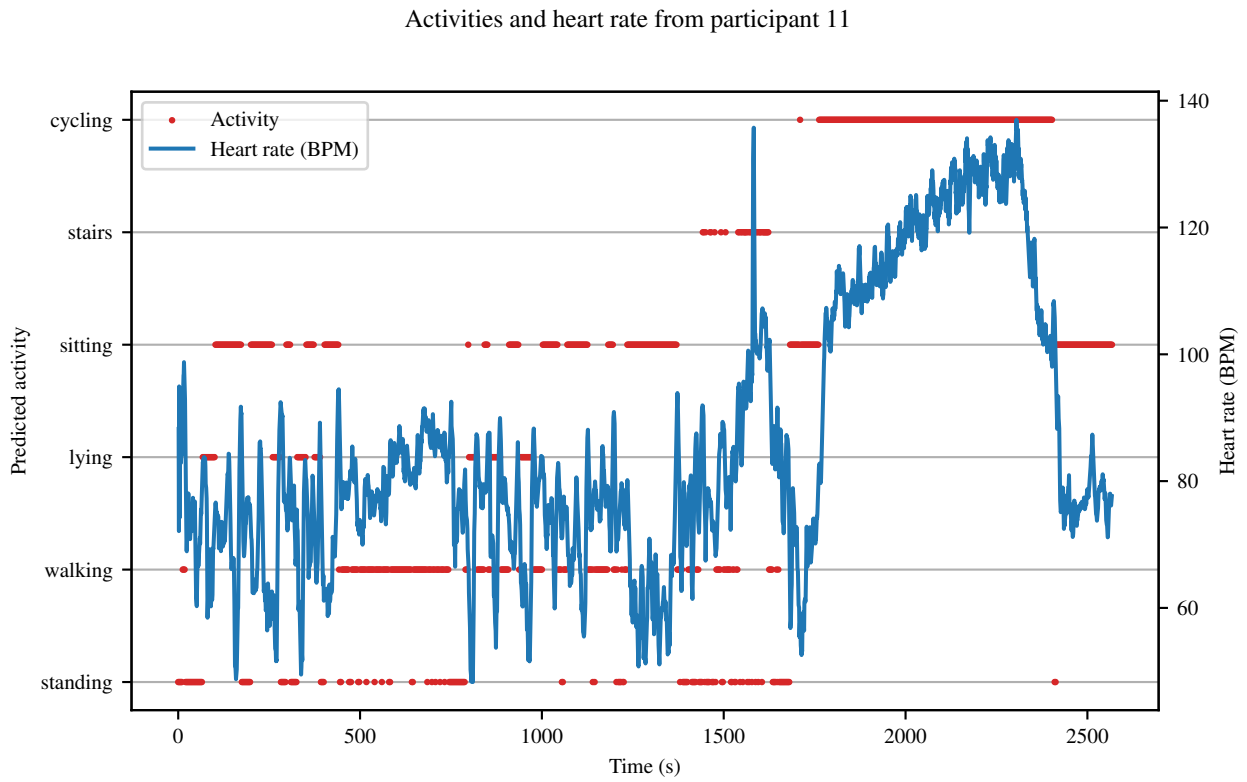

Figure 6. ECG recording overlaid on the predicted activities for experiment participant 11.

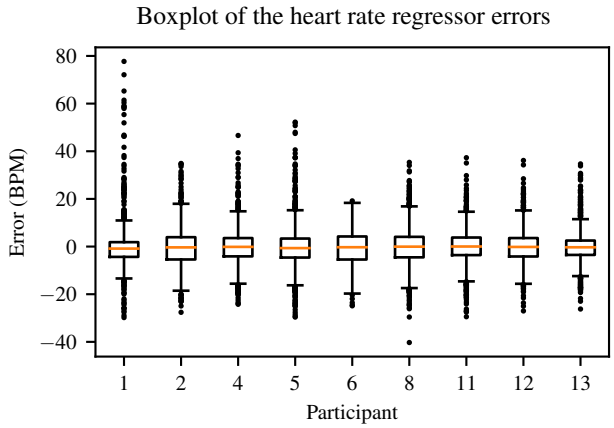

Figure 7. Boxplot showing the error in the output from the heart rate regressor model using predicted activity labels. The orange line shows the median, the box goes from the 25th to the 75th percentile and the black horizontal lines show the highest and lowest values excluding outliers. Values are classified as outliers if they are more than 1.5 times the interquartile range (IQR) removed from the median. IQR is defined as the distance between the 25th to the 75th percentile. Outliers are indicated in the plot by fliers—i.e., the black dots.

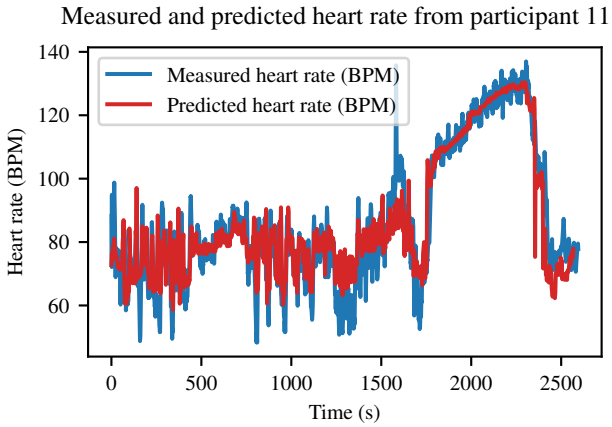

Figure 8. Predicted heart rate from the heart rate regressor model using predicted activity labels, overlaid on the measured heart rate for participant 11. The predicted heart rate stops slightly earlier than the measured heartrate, as the activity labels stopped earlier than the ECG recording.

of 86% confirmed the correlation between the input variables and heart rate: the score means that 86% of the variation in the predicted heart rate can be explained by the input variables.

Finally, Fig. 8 shows the predicted heart rate overlaid on the measured heart rate. As can be seen, the predicted heart rate follows the general trend of the measured heart rate. This is also reflected by the median absolute error value of 3.74 BPM (min 0 BPM, max 37.3 BPM), which means that the predicted BPM value is generally close to the measured BPM value.

IV. Discussion

The goal of this study was to find out whether Machine Learning could be used to create a data-driven model of the heart rate based on the output of a HAR system. We showed that it is possible to use a combination of a SVM classifier

and k-NN regressor to build a system that can predict the heart rate of healthy participants corresponding to a certain activity with reasonable accuracy. This was the case for a proof-of-concept using only a limited amount of data (approximately 30 minutes of training data per experiment participant).

A. Interpretation of results

1) Human Activity Recognition: As visible in the confusion matrices (Fig. 4 and 5), most activities are well distinguishable. Most confusion was between standing (median f1-score: 0.79) and walking (median f1-score: 0.80). The confusion matrixes show that most misclassifications for walking are standing, and vice versa. This is within expectation, as the transition between walking slowly and standing still is often rather vague. Because activity intensity is also considered for heart rate prediction, this confusion does not lead to large errors. The hip sensor gives slightly better results than the upper leg sensor for sitting and lying down in most cases. However, for this sensor the positioning is more important than it is for the upper leg sensor. If the hip sensor is mounted higher up on the hip, there is less rotation relative to the chest sensor. This makes distinguishing between sitting and standing more difficult. Additionally, movement of the legs is better captured by the upper leg sensor than by the hip sensor. This is particularly beneficial for distinguishing between walking and standing, and for recognizing walking up or down stairs. Performance for walking up and down stairs was often not satisfactory (median f1-score: 0.64). This was caused by that activity being similar to normal walking and because of low labelling accuracy. The latter was caused by the fact that the stairs were outside of view of the cameras and labelling was done by listening to the audio recording. The stairs that were used are also only four steps. However, this is not necessarily a limitation, as in a hospital setting patients often train by stepping up and down only a single step. A limitation is however that dominant leg was not considered. Someone's dominant leg might not necessarily be the same as their dominant arm, and this could have an effect on how well walking stairs is recognized.

Even though the experiment was conducted in a lab resembling a normal living space and participants were given freedom in how activities were performed, the experiment was still somewhat artificial. To illustrate, one of the participants noted 'you are very aware of what you are doing'. Additionally, because the eHealth House is a new environment for the participants, they might not perform activities the same way as they would in their own home.

2) Heart rate regressor: Fig. 8 shows that the regressor mostly smooths out variations in heart rate. It captures duration effects for longer activities—i.e., walking and cycling, generally quite well. Because the regressor takes into account the length of the current activity and the duration of cycling is relatively long, making a prediction for in-between points is comparatively straightforward. Transient spikes in heart rate are often not predicted. A possible reason for this is that there is limited training data available for such situations, particularly while walking stairs. Another point of interest is the very sudden dip in predicted heart rate visible in Fig. 8. Here the participant stopped cycling for a short period, which caused the HAR classifier to predict the activity 'sitting' at that point. Because the regressor does not consider the previous activity in predicting a heart rate, it assumes the heart rate immediately drops to a low heart rate, as most of the training data shows sitting associated with low heart rates. The plot might appear to show a larger maximum error than Table V. This is because the measured heart rate is plotted for every instant, whereas the regressor only considers the average heart rate during a single five-second data window.

B. Limitations

The experiment was relatively short and does therefore not capture long-term influences on heart rate, such as time of day. The cycling part was also done in one continuous run. Therefore, the effect of fatigue on the heart rate for different activities is not captured. While the regressor does take into account the length of the current activity, it does not consider the previous activity. A future model might take this information into account, though it is likely that more training data is required to cover all situations. The relatively small population of young healthy participants might also not be representative for a hospital setting, as patients have different health conditions and perform vastly different movements [64]. Care should therefore be taken to generalize the conclusions from this research to a hospital setting. Another limitation is that the current sensor setup is not comfortable to wear for longer periods of time. An integrated device, designed with wearability in mind, should be developed for clinical use. This device should also be able to extract more information form the ECG, such as arrhythmias. The current beat detection algorithm is not suitable for clinical use, as it is tuned for healthy people and the BioHarness. Additionally, the noise detection algorithm using the peak counter would be dangerous in a clinical setting, as periods with many peaks could also indicate serious problems—e.g., atrial fibrillation.

C. Future work

Using a photoplethysmogram (PPG) device to measure heart rate makes the system more comfortable to wear. There is some research that suggests that a Fitbit has comparable reliability as the BioHarness for heart rate measurements [65]. However, PPG is also considerably more sensitive to movement artefacts than ECG.

Better machine learning methods such as deep learning and neural networks could potentially be employed to create more accurate models for HAR and heart rate prediction. Additionally, a general model might be created that is applicable for many people instead of just one. Such a model could be based on relative heart rate instead of absolute heart rate. It is likely that a more involved experiment is needed for such a model, with more participants, also in a clinical setting. Special care should be taken that such a model will only be trained with typical heart rate data, and not with data that should lead to alarms, or at least that data should be treated as such. Creating a general model for heart rate

prediction is a larger challenge than creating a general model for HAR. This is due to the fact that the speed with which the heart rate changes is greatly dependent on physical condition, whereas the general movements people perform during a certain activity are similar.

It could be interesting to expand the system to also consider Heart Rate Variability (HRV) metrics, as they are an important indicator of cardiac health [66], [67]. However, a PPG device is not suitable for HRV analysis, as the pulse velocity through the blood vessels is not constant [29].

V. Conclusion

This work has shown that a HAR system can be implemented with simple sensors and classical machine learning techniques, while still achieving high accuracy (median 87%, min 82%, max 92%). This HAR system can then be used to feed a data-driven regressor model that predicts the heart rate for a certain activity with a certain intensity and duration with a median absolute error of median 3.82 BPM, minimum 2.94 BPM and maximum 4.79 BPM. This shows there is potential in using a combined HAR/heart rate prediction system to assist in continuous automatic heart rate monitoring, as the system can use the predicted heart rate to decide whether a measured heart rate is normal or cause for alarm.

Acknowledgement

The author would like to thank all experiment participants for freeing up some of their time to take part in the experiment. Additionally, he would like to thank Marcel Weusthof and Frodo Muijzer for their advice and help with the various sensors. Finally, he wants to thank the committee for their thorough and insightful comments and help. He is grateful in particular for Ying, who kept pushing to get this paper from 'good enough' to something I am proud of.

References

[1] S. C. Gandevia and D. K. McKenzie, "Respiratory rate: The neglected vital sign," *Medical Journal of Australia*, vol. 189, no. 9, p. 532, 2008.

[2] J. Ludikhuize, S. M. Smorenburg, S. E. de Rooij, and E. de Jonge, "Identification of deteriorating patients on general wards; measurement of vital parameters and potential effectiveness of the Modified Early Warning Score," *Journal of Critical Care*, vol. 27, no. 4, pp. 424.e7–424.e13, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.jcrc.2012.01.003

[3] S. C. Verrillo, M. Cvach, K. W. Hudson, and B. D. Winters, "Using Continuous Vital Sign Monitoring to Detect Early Deterioration in Adult Postoperative Inpatients," *Journal of Nursing Care Quality*, vol. 34, no. 2, pp. 107–113, 2019.

[4] S. R. Steinhubl, D. Feye, A. C. Levine, C. Conkright, S. W. Wegerich, and G. Conkright, "Validation of a portable, deployable system for continuous vital sign monitoring using a multiparametric wearable sensor and personalised analytics in an Ebola treatment centre," *BMJ Global Health*, vol. 1, no. 1, 2016.

[5] Nightingale, "Nightingale Public Final Report," no. October, pp. 1–8, 2021.

[6] T. Watkins, L. Whisman, and P. Booker, "Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit," *Journal of Clinical Nursing*, vol. 25, no. 1-2, pp. 278–281, 2016.

[7] R. G. Mendes, R. P. Simões, F. D. S. Costa, C. B. F. Pantoni, L. Di Thommazo-Luporini, S. Luzzi, O. Amaral-Neto, R. Arena, A. M. Catai, and A. Borghi-Silva, "Is applying the same exercise-based inpatient program to normal and reduced left ventricular function patients the best strategy after coronary surgery A focus on autonomic cardiac response," *Disability and Rehabilitation*, vol. 36, no. 2, pp. 155–162, 2014.

[8] L. Anderson, N. Oldridge, D. R. Thompson, A. D. Zwisler, K. Rees, N. Martin, and R. S. Taylor, "Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease Cochrane Systematic Review and Meta-Analysis," *Journal of the American College of Cardiology*, vol. 67, no. 1, pp. 1–12, 2016.

[9] M. Lamotte and S. Chimenti, "Resistive Training and Hemodynamics in Cardiac Rehabilitation," *Card Pulm Rehabil*, vol. 1, no. 2, pp. 1–4, 2017. [Online]. Available: https://www.omicsonline.org/open-access/resistive-training-and-hemodynamics-in-cardiac-rehabilitation.pdf

[10] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, Z. Yu, and S. Member, "Sensor-Based Activity Recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 42, no. 6, pp. 790–808, 2012.

[11] L. Chen and I. Khalil, "Activity Recognition: Approaches, Practices and Trends," in *Activity Recognition in Pervasive Intelligent Environments*, 2011, pp. 1–31.

[12] A. Davoudi, K. R. Malhotra, B. Shickel, S. Siegel, S. Williams, M. Ruppert, E. Bihorac, T. Ozrazgat-Baslanti, P. J. Tighe, A. Bihorac, and P. Rashidi, "Intelligent ICU for Autonomous Patient Monitoring Using Pervasive Sensing and Deep Learning," *Scientific Reports*, vol. 9, no. 1, pp. 1–13, 2019. [Online]. Available: http://dx.doi.org/10.1038/s41598-019-44004-w

[13] A. J. Ma, N. Rawat, A. Reiter, C. Shrock, A. Zhan, A. Stone, A. Rabiee, S. Griffin, D. M. Needham, and S. Saria, "Measuring Patient Mobiliy in the ICU Using a Novel Noninvasive Sensor," *Critical Care Medicine*, vol. 45, no. 4, pp. 630–636, 2017.

[14] E. Chou, M. Tan, C. Zou, M. Guo, A. Haque, A. Milstein, and L. Fei-Fei, "Privacy-Preserving Action Recognition for Smart Hospitals using Low-Resolution Depth Images," pp. 1–6, 2018. [Online]. Available: http://arxiv.org/abs/1811.09950

[15] P. H. Veltink, H. B. Bussmann, W. De Vries, W. L. Martens, and R. C. Van Lummel, "Detection of static and dynamic activities using uniaxial accelerometers," *IEEE Transactions on Rehabilitation Engineering*, vol. 4, no. 4, pp. 375–385, 1996.

[16] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3001, pp. 1–17, 2004.

[17] Ó. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[18] N. Ravi, H. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from Accelerometer Data," *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence*, vol. Volume 3, pp. 1541–1546, 2005.

[19] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, no. 2018, pp. 698–703, 2019. [Online]. Available: https://doi.org/10.1016/j.procs.2019.08.100

[20] M. Janidarmian, A. R. Fekr, K. Radecka, and Z. Zilic, "A comprehensive analysis on wearable acceleration sensors in human activity recognition," *Sensors (Switzerland)*, vol. 17, no. 3, 2017.

[21] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "A survey of online activity recognition using mobile phones," *Sensors (Switzerland)*, vol. 15, no. 1, pp. 2059–2085, 2015.

[22] J. Suto, S. Oniga, C. Lung, and I. Orha, "Comparison of offline and real-time human activity recognition results using machine learning techniques," *Neural Computing and Applications*, vol. 32, no. 20, pp. 15 673–15 686, 2020. [Online]. Available: https://doi.org/10.1007/s00521-018-3437-x

[23] F. R. Halfwerk, J. H. van Haaren, R. Klaassen, R. W. van Delden, P. H. Veltink, and J. G. Grandjean, "Objective quantification of in-hospital patient mobilization after cardiac surgery using accelerometers: Selection, use, and analysis," *Sensors*, vol. 21, no. 6, pp. 1–18, 2021.

[24] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring motor fluctuations in patients with parkinsons disease using wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 864–873, 2009.

[25] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H. P. Tan, "Deep activity recognition models with triaxial accelerometers," *AAAI Workshop - Technical Report*, vol. WS-16-01 -, pp. 8–13, 2016.

[26] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019. [Online]. Available: https://doi.org/10.1016/j.patrec.2018.02.010

[27] X. Zhou, W. Liang, K. I. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-Learning-Enhanced Human Activity Recognition for Internet of

Healthcare Things," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6429–6438, jul 2020.

[28] K. Xia, J. Huang, and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," *IEEE Access*, vol. 8, pp. 56 855–56 866, 2020.

[29] F. Shaffer, R. McCraty, and C. L. Zerr, "A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability," *Frontiers in Psychology*, vol. 5, no. September, pp. 1–19, 2014.

[30] I. Cygankiewicz and W. Zareba, *Heart rate variability*, 1st ed. Elsevier B.V., 2013, vol. 117. [Online]. Available: http://dx.doi.org/10.1016/B978-0-444-53491-0.00031-6

[31] R. Perini and A. Veicsteinas, "Heart rate variability and autonomic activity at rest and during exercise in various physiological conditions," *European Journal of Applied Physiology*, vol. 90, no. 3-4, pp. 317–325, 2003.

[32] C. Schubert, M. Lambertz, R. A. Nelesen, W. Bardwell, J. B. Choi, and J. E. Dimsdale, "Effects of stress on heart rate complexity-A comparison between short-term and chronic stress," *Biological Psychology*, vol. 80, no. 3, pp. 325–332, 2009.

[33] M. Jacobsen, T. A. Dembek, G. Kobbe, P. W. Gaidzik, and L. Heinemann, "Noninvasive Continuous Monitoring of Vital Signs With Wearables: Fit for Medical Use?" *Journal of Diabetes Science and Technology*, vol. 15, no. 1, pp. 34–43, 2020.

[34] Y. Eddahchouri, R. V. Peelen, M. Koeneman, H. R. Touw, H. van Goor, and S. J. Bredie, "Effect of continuous wireless vital sign monitoring on unplanned ICU admissions and rapid response team calls: a before-and-after study," *British Journal of Anaesthesia*, no. January, pp. 1–7, 2022. [Online]. Available: https://doi.org/10.1016/j.bja.2022.01.036

[35] T. M. Cheng, A. V. Savkin, B. G. Celler, S. W. Su, and L. Wang, "Nonlinear modeling and control of human heart rate response during exercise with various work load intensities," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 11, pp. 2499–2508, 2008.

[36] M. S. Olufsen and J. T. Ottesen, "A practical approach to parameter estimation applied to model predicting heart rate regulation," *Journal of Mathematical Biology*, vol. 67, no. 1, pp. 39–68, 2013.

[37] M. Alber, A. Buganza Tepole, W. R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W. W. Lytton, P. Perdikaris, L. Petzold, and E. Kuhl, "Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences," *npj Digital Medicine*, vol. 2, no. 1, 2019. [Online]. Available: http://dx.doi.org/10.1038/s41746-019-0193-y

[38] M. Li, V. Rozgić, G. Thatte, S. Lee, A. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan, "Multimodal physical activity recognition by fusing temporal and cepstral information," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, pp. 369–380, 2010.

[39] R. Jia and B. Liu, "Human daily activity recognition by fusing accelerometer and multi-lead ECG data," *2013 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC 2013*, pp. 9–12, 2013.

[40] J. Liu, J. Chen, H. Jiang, W. Jia, Q. Lin, and Z. Wang, "Activity Recognition in Wearable ECG Monitoring Aided by Accelerometer Data," *Proceedings - IEEE International Symposium on Circuits and Systems*, vol. 2018-May, pp. 3–6, 2018.

[41] A. Stojmenski, M. Gusev, L. Poposka, and M. Vavlukis, "Heart rate variability of 30-minute ECG measurements and correlation to glucose levels," *2020 28th Telecommunications Forum, TELFOR 2020 - Proceedings*, pp. 1–4, 2020.

[42] R. Amirhossein, S. Amirhossein, and M. Maryam, "Sleep Stage Classification Based on ECG-Derived Respiration and Heart Rate Variability of Single-Lead ECG Signal," in *2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME)*, 2019, pp. 158–163.

[43] M. A. Quiroz-Juarez, O. Jimenez-Ramirez, R. Vazquez-Medina, E. Ryzhii, M. Ryzhii, and J. L. Aragon, "Cardiac Conduction Model for Generating 12 Lead ECG Signals with Realistic Heart Rate Dynamics," *IEEE Transactions on Nanobioscience*, vol. 17, no. 4, pp. 525–532, 2018.

[44] T. Pawar, N. S. Anantakrishnan, S. Chaudhuri, and S. P. Duttagupta, "Transition detection in body movement activities for wearable ECG," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 6, pp. 1149–1152, 2007.

[45] F. T. Sun, C. Kuo, and M. Griss, "PEAR: Power efficiency through activity recognition (for ECG-based sensing)," *2011 5th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2011*, pp. 115–122, 2011.

[46] F. T. Z. Iqbal, K. A. Sidek, N. A. Noah, and T. S. Gunawan, "A comparative analysis of QRS and cardioid graph based ECG biometric recognition in different physiological conditions," *2014 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2014*, no. November, pp. 25–27, 2015.

[47] K. A. Sidek, I. Khalil, and M. Smolen, "ECG biometric recognition in different physiological conditions using robust normalized QRS complexes," *Computing in Cardiology*, vol. 39, pp. 97–100, 2012.

[48] H. Verma, D. Paul, S. R. Bathula, S. Sinha, and S. Kumar, "Human Activity Recognition with Wearable Biomedical Sensors in Cyber Physical Systems," *INDICON 2018 - 15th IEEE India Council International Conference*, 2018.

[49] L. S. Pescatello, R. Arena, D. Riebe, and P. D. Thompson, "Cycle ergometer testing," *American College of Sports Medicine*, vol. 9th, pp. 218–221, 2014.

[50] M. E. Haveman, M. C. van Rossum, R. M. Vaseur, C. van der Riet, R. C. Schuurmann, H. J. Hermens, J. P. P. de Vries, and M. Tabak, "Continuous Monitoring of Vital Signs With Wearable Sensors During Daily Life Activities: Validation Study," *JMIR Formative Research*, vol. 6, no. 1, pp. 1–16, 2022.

[51] Xsens, "MTw Awinda User Manual MTw," no. May, 2018.

[52] Zephyr, "BioHarness 3.0 User Manual," no. September, 2012.

[53] S. T. Boerema, L. van Velsen, L. Schaake, T. M. Tönis, and H. J. Hermens, "Optimal sensor placement for measuring physical activity with a 3D accelerometer," *Sensors (Switzerland)*, vol. 14, no. 2, pp. 3188–3206, 2014.

[54] J. Reback, jbrockmendel, W. McKinney, J. V. den Bossche, T. Augspurger, P. Cloud, S. Hawkins, M. Roeschke, gfyoung, Sinhrks, A. Klein, P. Hoefler, T. Petersen, J. Tratner, C. She, W. Ayd, S. Naveh, M. Garcia, J. Darbyshire, J. Schendel, R. Shadrach, A. Hayden, D. Saxton, M. E. Gorelli, F. Li, M. Zeitlin, V. Jancauskas, A. McMaster, P. Battiston, and S. Seabold, "pandas-dev/pandas: Pandas 1.4.0," Jan. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.5893288

[55] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, Stéfan van der Walt and Jarrod Millman, Eds., 2010, pp. 56 – 61.

[56] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[57] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[58] S.-L. Developers, "6.3. preprocessing data," 2022. [Online]. Available: https://scikit-learn.org/stable/modules/preprocessing.html#standardization-or-mean-removal-and-variance-scaling

[59] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. BME-32, no. 3, pp. 230–236, 1985.

[60] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[61] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography." *European heart journal*, vol. 13, no. 9, pp. 1164–1172, sep 1992.

[62] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[63] G. James, D. Witten, T. Hastie, and R. Tibshirani, *Linear Regression*. New York, NY: Springer New York, 2013, pp. 59–126. [Online]. Available: https://doi.org/10.1007/978-1-4614-7138-7_3

[64] C. Hodgson, D. Needham, K. Haines, M. Bailey, A. Ward, M. Harrold, P. Young, J. Zanni, H. Buhr, A. Higgins, J. Presneill, and S. Berney,

"Feasibility and inter-rater reliability of the ICU Mobility Scale," *Heart and Lung: Journal of Acute and Critical Care*, vol. 43, no. 1, pp. 19–24, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.hrtlng.2013.11.003

[65] G. Nazari, J. C. MacDermid, K. E. Sinden, J. Richardson, and A. Tang, "Reliability of Zephyr Bioharness and Fitbit Charge Measures of Heart Rate and Activity at Rest, During the Modified Canadian Aerobic Fitness Test, and Recovery," *Journal ofStrength and Conditioning Research*, vol. 33, no. 2, 2019.

[66] R. Sassi, S. Cerutti, F. Lombardi, M. Malik, H. V. Huikuri, C. K. Peng, G. Schmidt, and Y. Yamamoto, "Advances in heart rate variability signal analysis: Joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society," *Europace*, vol. 17, no. 9, pp. 1341–1353, 2015.

[67] T. F. of the European Society of Cardiology the North American Society of Pacing Electrophysiology, "Heart Rate Variability," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

APPENDIX A
EXPERIMENT PROTOCOL

A summary of the experiment protocol is found in Table VII.

Table VII
EXPERIMENT PROTOCOL

Activity	Duration	Total duration of part
Controlled activities		∼10 min (some extra changing time for changing between activities)
Standing	30 s	
Lying	30 s	
Sitting (bed)	30 s	
Sitting (chair)	30 s	
Standing	20 s	
Sitting (chair)	20 s	
Sitting (bed)	20 s	
Lying	20 s	
Standing	10 s	
Sitting (bed)	10 s	
Standing	10 s	
Lying	10 s	
Sitting (chair)	10 s	
Lying	10 s	
Standing	10 s	
Sitting (chair)	10 s	
Walking (slow)	20 s	
Walking (medium)	20 s	
Walking (fast)	20 s	
Free activities		About 15–20 minutes
Walk from bedroom to kitchen, stop, turn around and walk back (3x)		
Fetch a cup from the kitchen and bring it back (3x)		
Sit in a chair, walk to the other room and lie down (3x)		
Various other activities, tbd during experiment		
Walking up and down stairs (3x)		
Cycling – 50 RPM		∼10 minutes
Warmup	3 min	
Test – about 140 bpm	6 min	
Cooling down	1 min	

APPENDIX B
ACTIVITY LOGGER

To annotate the sensor data, a custom activity logger application was written in Python. A screenshot of the application can be found in Fig. 9. It works by filling in the starting time timestamp, choosing a file to write the data to (or creating a new file), pressing start (both in the logger and the experiment recording player) and then pressing the buttons corresponding to the activity currently performed by the experiment participant. When done, the application is quit by pressing the exit button. All data is saved in a .csv format. Sitting (bed) and sitting (chair) are both interpreted as sitting. Walking, stopping and turning are all interpreted as walking. Walking upstairs and walking downstairs are both interpreted as walking stairs, and finally all types of cycling are interpreted as cycling. In case there were instances of 'other activity', they were manually interpreted.

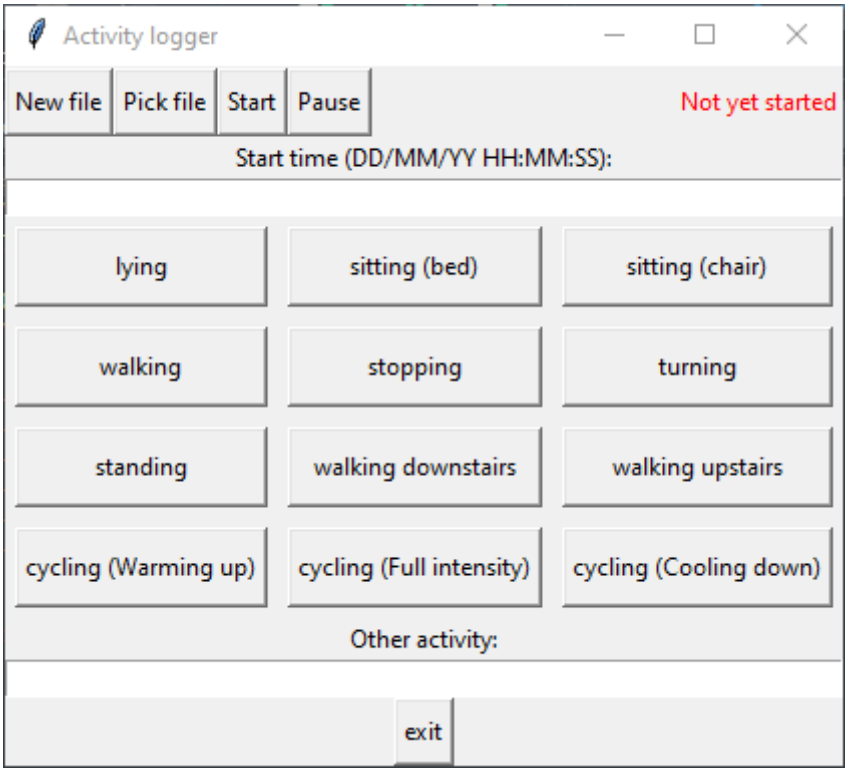

Figure 9. Screenshot of the activity logger app.

APPENDIX C
FULL RESULTS

The full HAR results per participant, per activity can be found in Table VIII. Also present in this table are average heart rates during each activity and the standard deviation of the heart rate during each activity.

Table VIII
PER-ACTIVITY PERFORMANCE METRICS FOR HUMAN ACTIVITY RECOGNITION AND MEAN HEART RATE AND STANDARD DEVIATION OF THE HEART RATE DURING EACH ACTIVITY.

Participant	Activity	Precision	Recall	f1-score	Support	Mean Heart Rate (in BPM)	Standard Deviation of Heart Rate (in BPM)
1	standing	0.73	0.72	0.72	205	61.5	9.20
	walking	0.63	0.68	0.65	210	64.3	14.5
	cycling	0.98	0.97	0.98	258	102.5	16.9
	lying	0.90	0.77	0.83	73	54.4	8.01
	sitting	0.88	0.91	0.89	257	54.4	6.23
	stairs	0.68	0.56	0.61	41	73.9	23.2
2	standing	0.78	0.80	0.79	203	76.8	12.7
	walking	0.74	0.74	0.74	167	82.7	9.51
	cycling	0.99	0.99	0.99	223	108.0	13.3
	lying	0.98	0.82	0.90	57	74.5	10.8
	sitting	0.90	0.91	0.90	244	71.2	9.14
	stairs	0.65	0.63	0.64	38	94.8	8.27
4	standing	0.75	0.80	0.77	196	92.3	10.2
	walking	0.69	0.71	0.70	134	90.6	10.4
	cycling	0.98	0.98	0.98	236	119.1	12.9
	lying	0.88	0.80	0.84	75	74.2	8.66
	sitting	0.94	0.92	0.93	270	78.4	9.55
	stairs	0.67	0.61	0.64	49	105.9	9.72
5	standing	0.83	0.83	0.83	218	73.0	9.84
	walking	0.78	0.83	0.80	210	77.8	11.4
	cycling	1.00	0.98	0.99	278	106.7	12.9
	lying	0.97	0.96	0.96	99	62.5	9.07
	sitting	0.98	0.97	0.98	204	63.6	10.2
	stairs	0.71	0.62	0.66	55	86.3	11.1
6	standing	0.79	0.82	0.80	184	98.3	7.02
	walking	0.82	0.82	0.82	203	99.1	7.24
	cycling	0.98	0.98	0.98	43	109.8	10.5
	lying	1.00	0.98	0.99	60	83.9	9.89
	sitting	0.95	0.94	0.95	175	85.1	8.58
	stairs	0.89	0.84	0.86	49	108.6	6.82
8	standing	0.86	0.81	0.83	185	93.2	10.0
	walking	0.79	0.85	0.82	215	96.5	7.70
	cycling	1.00	1.00	1.00	263	123.9	8.73
	lying	0.99	0.86	0.92	81	75.0	9.70
	sitting	0.96	0.98	0.97	432	79.6	10.7
	stairs	0.81	0.75	0.78	59	104.4	7.06

Table VIII
(CONTINUED)

Participant	Activity	Precision	Recall	f1-score	Support	Mean Heart Rate (in BPM)	Standard Deviation of Heart Rate (in BPM)
11	standing	0.69	0.67	0.68	157	78.9	8.39
	walking	0.79	0.81	0.80	210	80.2	5.46
	cycling	0.98	0.98	0.98	258	117.8	11.7
	lying	0.99	0.93	0.96	74	69.1	11.7
	sitting	0.90	0.93	0.92	295	70.9	9.38
	stairs	0.75	0.53	0.62	34	94.5	11.9
12	standing	0.83	0.77	0.80	143	83.0	9.31
	walking	0.85	0.86	0.86	213	91.4	6.97
	cycling	0.99	0.99	0.99	263	123.4	13.4
	lying	0.96	0.96	0.96	67	79.2	10.8
	sitting	0.92	0.96	0.94	218	79.5	11.2
	stairs	0.90	0.86	0.89	42	106.1	8.97
13	standing	0.66	0.78	0.71	161	81.1	13.4
	walking	0.81	0.77	0.79	166	77.3	6.74
	cycling	1.00	0.99	0.99	280	115.9	15.0
	lying	1.00	0.71	0.83	31	64.9	6.29
	sitting	0.96	0.91	0.94	150	65.2	7.07
	stairs	0.64	0.60	0.62	35	97.1	9.93
Aggregate	standing	0.78	0.77	0.78	1652	81.7	15.1
	walking	0.75	0.78	0.76	1728	84.4	14.2
	cycling	0.98	0.98	0.98	2102	114.7	15.1
	lying	0.94	0.93	0.94	617	70.5	13.0
	sitting	0.94	0.93	0.94	2245	72.5	13.1
	stairs	0.68	0.61	0.64	402	97.3	15.7

APPENDIX D
FEATURE SELECTION

The following two pages contain the tables used to determine which features were used.

Expected signals

IMUs

X/Y/Z acceleration gives the same information as current angle, as the acceleration depends on gravity (=9.81 m/s^2).

Lying

LOCATION	X	Y	Z	GYRO	DYNAMICS
CHEST	0	0	9.81	0	Low
UPPER ARM	0	9.81	0	0	Low
LOWER ARM/WRIST	0	Depends on wrist rotation	Depends on wrist rotation	0	Low
HIP	~5	0	~5	0	Low
UPPER LEG	0	0	9.81	0	Low

Sitting

LOCATION	X	Y	Z	GYRO	DYNAMICS
CHEST	9.81	0	0-5	0	Low
UPPER ARM	9.81	0	0-5	0	Low
LOWER ARM/WRIST	Low	~9.81	Medium	0	Depends on if e.g. typing
HIP	5-9.81	9.81	0	0	Low
UPPER LEG	0-5	0	9.81	0	Low

Standing

LOCATION	X	Y	Z	GYRO	DYNAMICS
CHEST	9.81	0	0-5 (depends on bending)	0, unless turning around	Low
UPPER ARM	9.81	0	~5	0	Low
LOWER ARM/WRIST	Depends on if e.g. carrying something				
HIP	0	9.81	0	0	Low
UPPER LEG	9.81	0	0	0	Low

Walking

LOCATION	X	Y	Z	GYRO	DYNAMICS
CHEST	9.81 +-5	0 +- 3	2.5 +- 2	Mostly X, +-50	Periodic, mostly in X
UPPER ARM	9.81 +- 4	0 +- 4	5 +- 2	Mostly X, Z, +- 75	Arm swings once every two steps
LOWER ARM/WRIST	2.5 +- 2.5	9.81 +- 3	-2 +- 2	Low activity	Period, but depends on if carrying smth
HIP	0 +- 6	9.81 +- 5	0 +- 3	All channels, +- 75	Periodic
UPPER LEG	9.81 +- 8	0 +- 7	0 +- 10	X +- 350, Y +- 200	Periodic

Walking stairs

LOCATION	X	Y	Z	GYRO	DYNAMICS
CHEST	9.81 +- 6	0 +- 3	1 +- 3	All +- 50	Periodic, slower than walking
UPPER ARM	9.81 +- 5	0 +- 3	3 +- 3	X +- 100, Y, Z +- 40	Not as periodic as walking
LOWER ARM/WRIST	9.81 +- 7	2.5 +- 3	0 +- 2	Mostly X, Z +- 150	Periodic mostly in Z
HIP	0 +- 4	9.81 -5 +10	2.5 +- 2	All +- 75	Periodic
UPPER LEG	9.81 +- 8	0 +- 8	0 +15 -10	X, Y +- 200, Z +- 75	Periodic

Cycling

LOCATION	X	Y	Z	GYRO	DYNAMICS
CHEST	9.81	0	0	X +- 30, Y +- 10, Z +- 20	Gyro: periodic, X and Y antiphase. Acc: medium
UPPER ARM	7 +- 1	7 +- 1.5	2.5 +- 2.5	X +- 30, Y +- 15, Z +- 10	Periodic, Acc mostly in Z
LOWER ARM/WRIST	2	3	9.81	All +- 15	Gyro periodic
HIP	3 +- 1	9 +- 1	2 +- 2	X +- 40, Y +- 20, Z +- 20	Acc Z, Gyro periodic
UPPER LEG	6 +- 6	-3 +- 4	6 +- 2	X +- 60, Y +- 130 sinusoid, Z +- 20	Periodic

Conclusions

Chest and hip or upper leg should discriminate between all activities. Upper arm does not add that much, similar info as chest. Lower arm/wrist very much depends on if the person is doing something with their hand.